Multitaper Spectral Processing for Acoustic Detection of UUVs

Milo Prisbrey[1*], Michael Teti[1*], Marko Orescanin[2], and Benjamin Migliori[1]

[1]Los Alamos National Laboratory, P.O. Box 1663, Los Alamos NM
[2]Naval Postgraduate School, 1 University Cir, Monterey

Benjamin Migliori, Los Alamos National Laboratory, ben.migliori@lanl.gov

Abstract: *The detection of unmanned undersea vehicles (UUVs) represents a prominent area of application for underwater acoustic methods. Successful vehicle tagging requires an understanding of the background noise, the vehicle parameters and configuration, and the acoustic propagation environment. Here, we present a novel non-parametric method for preprocessing hydrophone data for UUV detection based on multitaper analysis. Multitaper methods use multiple overlapping tapers, each an orthogonal function that is localized in both time and frequency, to compute part of a spectral estimate. The full estimate is the weighted sum of the individual taper estimates, where the weights are chosen to optimize the spectral resolution and for variance reduction. The multitaper method has been shown to be particularly effective in cases where the signal is contaminated with noise or other sources of interference, as it can suppress such sources more effectively than traditional methods. In this work, we show how to configure multitaper analysis for UUV detection, and show spectral comparisons with more traditional methods. Finally, we show how to use spectral estimates as inputs for neural networks, and describe suitable architectures that match the physics of the acoustic data. We demonstrate the efficacy of our methods on UUV data acquired in Lake Crescent, Washington. We show the the use of physics-informed features, on the form of multitaper spectral estimates, provide a computationally efficient and high performance data view on otherwise normally acquired hydrophone data. While we demonstrate our method for the desirable case of detection of UUV acoustics, this technique is expected to be applicable to shipping traffic, marine biome acoustics, and natural undersea phenomena.*

Keywords: *passive acoustics monitoring, audio classification, multitaper spectral analysis*

[*]Equal contribution

1. INTRODUCTION

The classification of acoustic objects is a daily experience in modern life; from smart assistants in miniature speakers to children's toys, acoustic events (glass breaking, voice commands) are constantly being analyzed. However, the ability to identify complex acoustic textures (irregular hums, vibrations, and drones) remains much more primitive [1].

In this paper, we demonstrate the use of spectral estimates over a physics-based sampling period to create useful feature projections for machine learning detection of acoustic textures. We focus on acoustic textures produced by the propulsion systems of small unmanned undersea vehicles (UUVs) in shallow-water environments. Our results use a technique common in neuroscience, called *multitaper spectral estimation* [2], to create high quality estimations of spectral content. We use coarse estimates of the cyclostationarity of the UUV propulsion system acoustic output (based on vessel speed and response time) to configure our spectral sample length. We show that even in the case of limited data, effective classifiers may be developed. This is primarily due to the nature of the screw-based propulsion system, in which the speed of the propeller, the bathymetry of the water, and the state of the UUV all interact to produce complex but spectrally-rich sounds.

The more common approach, timeseries classification, typically requires a known acoustic event (smart speaker wake word, for example) to trigger the sample period. For detection of a UUV, synchronization is usually physically infeasible. Spectral analysis is similar to bag-of-words classification in natural language processing, where the spectral amplitude (much like number of instances of each word in a document) is used to produce a vector in a high dimensional space (the vocabulary, or the Nyquist range in our case). Given a high enough dimensionality in the space, these feature vectors provide good discrimination between unlike sources. Frequency space classification has been used in a wide range of applications, from detection of ovarian cancer [3] to identification of music [4].

However, discrete representations result in non-ideal spectra due to numerical artifacts introduced through the FFT process. The discrete spectrum of a pure tone is lobe straddling the central frequency surrounded by side-lobes of decreasing power, commonly referred to as spectral-leakage [5]. Traditional techniques used to reduce spectral leakage windows the data (via a taper function) such that it has the same value and first derivative at the beginning and end, satisfying the infinite-repeatability assumption inherent in the FFT [6]. Though this process does reduce leakage, it does not fully solve the problem. Several taper functions are commonly used, including Hann, Hamming, and Tukey functions.

Multitaper spectral estimation, a non-parametric method for estimating the spectral density of a signal in the presence of noise contamination, uses multiple overlapping tapers to increase the amplitude of the main lobe and decrease the amplitude of the side lobes [2]. This method is particularly effective where the signal is contaminated with uncorrelated noise. The discrete prolate spheroidal sequences [7] are used to create the taper sequence. Each taper is constructed to have the maximum energy concentration within a given frequency band, while minimizing the energy outside of that band. This makes the tapers provide a good balance between frequency resolution and side-lobe energy. The catch (there is no such thing as a free lunch) is that this method assumes that the underlying phenomena is both cyclostationary and ergodic, i.e. that over the sampling windows the underlying "generator" will have repeated behavior and trend to a stable value [8]. This matches the physics of UUV propulsion systems, and results in highly effective machine learning (ML) classifiers for an otherwise challenging and small dataset of acoustic observations.

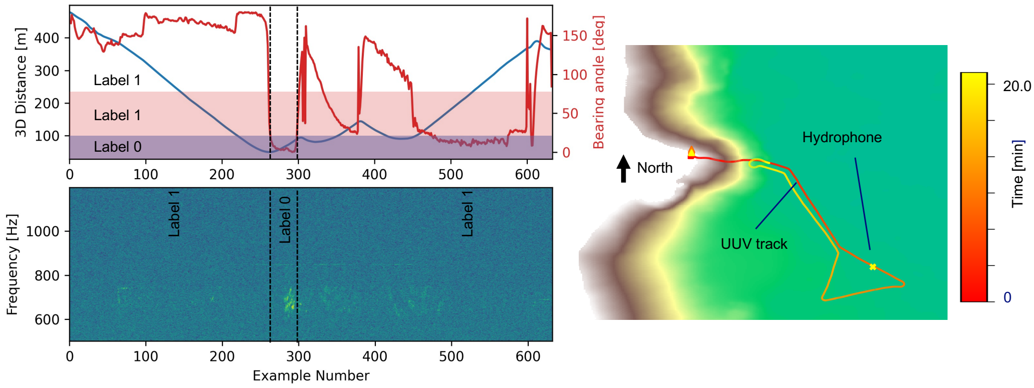

Figure 1: ***Correlated navigational and acoustic data.*** *The top panel shows the distance and bearing from the recording hydrophone to the UUV as a function of time. The bottom panel shows a multitaper spectrogram from the same hydrophone, in which the effects of thruster direction may be seen. Inset on the right is a rendering of the bathymetry data and navigational path.*

2. METHODS

2.1. ACOUSTIC RECORDINGS AND NAVIGATIONAL DATA

Acoustic recordings were taken in Lake Crescent, Washington, USA. Acoustimetrics Acousonde Model 3A hydrophones were placed on moorings in the water prior to experiment start. A small UUV was then deployed from the dock and followed a set course in the water (Figure 1). During the course of the deployment, the hydrophones were recording constantly. The total deployment lasted 26.3 minutes.

The acoustic recordings and GPS / navigational data from the UUV were synchronized using Python. The acoustic data was converted into a binary format. The data was then labelled based on the UUV's relative bearing and distance to the hydrophone, which was established from the navigational data acquired by the UUV (Figure 1).

2.2. MULTITAPER PROCESSING

For frequency-space classification, the acoustic recording of duration T_s was segmented into smaller signals of duration $T_c << T_s$. Each segment is then multiplied with a taper function and converted via Fast Fourier Transform (FFT). Multitaper spectral analysis differs from traditional techniques by averaging multiple FFT realizations of the same audio clip (T_c) together, resulting in a spectrum estimate $Z(f)$. This estimate consists of estimates from many different tapers, reducing variance in the resulting spectrum. $Z(f)$ calculated as

$$Z(f) = \frac{1}{L}\sum_{n=1}^{L} |FFT(Z(t)h_n(t))|^2, \tag{1}$$

where

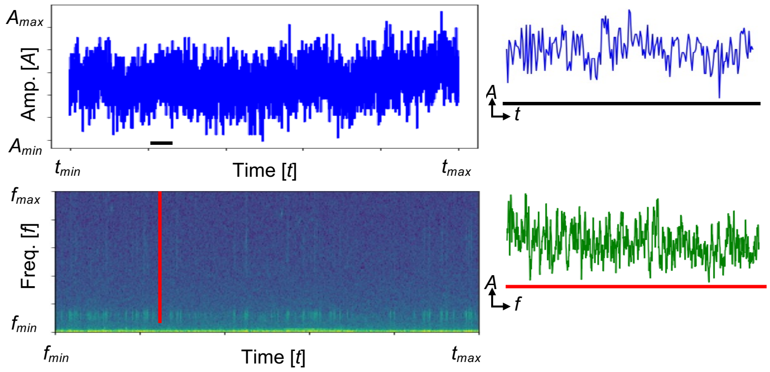

Figure 2: ***Timeseries analysis vs spectral analysis for machine learning.*** *The top panel shows a timeseries audio recording of a UUV, with a two-second window shown. This window is a typical ML data input. The black bar and corresponding inset shows a typical timeseries input for an ML classifier. The bottom panel shows a multitaper spectrogram. The vertical red bar and corresponding inset indicates the spectral example used as an input to the classifier.*

$$L = \max(1, \lfloor \frac{M}{f_s} * df - 1 \rfloor). \tag{2}$$

The value df (bandwidth in Hz) controls the resolvable frequencies in the spectral estimate. df should be chosen to be large enough such that unwanted noise is filtered out but small enough that important spectral dynamics are not obscured. M is the length (in number of points) of the audio clip being transformed, which controls the temporal resolution. Finally, $h_n(t)$ is the n^{th} Slepian (also widely known as discrete prolate spheriodal sequence), which belongs to a set of band-limited, orthogonal taper functions. Ideally, the choice of the value L is made by choosing df and M to reflect the spectral dynamics and known physics of the signal. While the argument can be made that the number of tapers must remain below the Shannon-Nyquist number[9], and $L > 1$ must also hold for the term multitaper to apply, a practical lower bound is less objective. Empirically, a value of L on the order of 10^1 provides the most benefit.

In the case of the UUV acoustic recordings, we utilized the known velocities of the vehicle (3-5 knots) and typical navigational controls (derived from the programmed mission path) to assess M. For df, we knew that we could expect Doppler shifts on the order of 2-10 Hz, and thus chose df to resolve these shifts.

2.3. DATA AUGMENTATION

With a df of 4, this resulted in a approximately 800 examples for training and testing of the UUV recordings. The multitaper spectral projection is “clean” enough to train a simple classifier, but we still needed to deal with the challenge of data imbalance. To better balance the data, we chose to use SMOTE (Synthetic Minority Oversampling Technique)[10]. SMOTE generates synthetic data by selecting samples from the minority class and computing the k-nearest neighbors for these samples. The minority class is over-sampled by taking each minority class sample centroid and introducing synthetic examples along the line segments joining the k minority class nearest neighbors. SMOTE generates synthetic samples that are similar to the existing minority class samples but are not identical to them. This helps to reduce overfitting and improve the generalization performance of machine learning models.

The original data set was significantly imbalanced (see Figure 1) and was SMOTEd to have a 70:30 class distribution for training. These classes were defined to leverage the limited size of the lake and the observed spectral power observed in the acoustic data.

2.4. MACHINE LEARNING (ML) ANALYSIS

Because there is no physical meaning to shared feature kernels across spectral frequencies, an MLP is a numerically efficient and simple method for fitting the data. Similarly, a 1D convolutional net could have been used; the critical point is to prevent weight sharing along the frequency dimension. The dataset (SMOTEd and inference) was divided partitioned into 90% training and 10% testing sets. A multilayer perceptron (MLP) with dropout was used to fit the data (depth three, dimensions 1000, 500, 100, with dropout at 10% between layers). L1/L2 regularization [11] was used to limit overfitting. The model was trained until convergence in Keras, with early stopping enabled.

3. RESULTS

Training on naively extracted timeseries data (two second window, 51,622 length), we were unable to get the model to converge. Using the multitaper spectral estimate, the model converged rapidly on single-GPU / laptop class machines. The resulting confusion matrix, for classes "UUV within 100 meters and angle of bearing 45 degrees" is shown in Figure 3. The performance is fully described in Table 1. Figure 3 represents the first successful application of multitaper spectral techniques to the UUV detection problem.

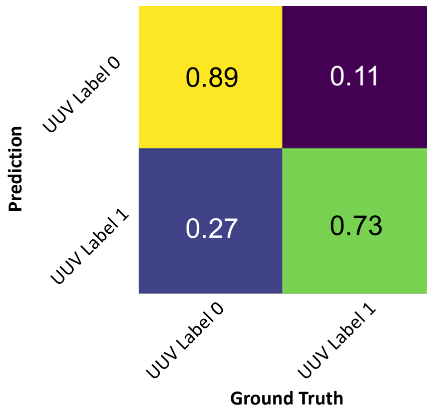

Figure 3: ***Classification performance.*** *The data was tested against two classes, "UUV within 100 meters and 45 degree angle of bearing (Class 0)", and "UUV outside 100 meters or not within 45 degrees angle (Class 1)".*

Table 1: Amount of Each Class in the Dataset

	Label 0	**Label 1**
Accuracy	0.89	0.73

4. DISCUSSION

During the operation of a typical unmanned undersea system, there are specific limits to the motion of that system. These limits can arise from a wide variety of sources, ranging from the physical limits of the device (power output, fin size) to environmental regulations (maximum emitted sound pressure level). By considering the system to be detected, *semi-engineered* parameters such as these can be used to make a coarse estimate of vessel maximum velocity and maximum rate of bearing change. Multitaper spectral acoustic classification uniquely leverages these parameters; configuring the spectral estimate is somewhere in between hand-engineered and data driven features. By choosing df and M to resolve the frequencies and timecourse of the dynamic features of vessel motion, previously challenging detection problems can be solved with less data and less compute. We consider this to be a mechanism for reducing the difficulty level of the problem faced by the classifier model itself; by knowing that useful information lives in the spectral domain, we assist the classifier in reaching that view. Linking this spectral domain to cleaner, less leaky multitaper estimates allows relatively small architectures to be successful with small amounts of data.

The techniques presented here will become especially useful as the explosion in consumer drones expands from the aerial type (UAVs) to ocean-going UUVs. For sanctuaries and ecologically preserved areas, there is currently no economically feasible way to detect UUVs that disregard the preservation rules. As in the aerial domain [12], especially in firefighting operations, stating the rules to pilots is not enough to keep unmanned systems out of restricted areas. As visual means are especially challenging for detection of UUVs, low-cost, low-complexity autonomous detection systems represent a potential solution for conservation efforts. It is worth noting that the models presented here are easily within the reach of portable consumer hardware such as modern laptops, and possibly implementable on edge computing devices such as Raspberry Pi single-board computers.

5. CONCLUSION

Here, we present a novel approach based on the use of physical parameter estimation to construct highly efficient spectral feaures for UUV detection in the aocustic domain. Our passive acoustic method results in highly selective UUV detection, under real-world conditions in a large freshwater lake. The selection of parameters relies on extremely coarse physical knowledge of the UUVs to be detected, and is used to generate data-driven feature representations. We show that simple ML architectures are then able to successfully classify the presence of absence of UUVs via omnidirectional hydrophones.

Our method represents a knew way of approaching shallow water acoustic soundscape classification, and has significant size, weight, and power advantages to naive timeseries methods. This method may be useful in the future for prevention of UUV driven fishing / line casting or other intrustions into ecologically protected regions.

REFERENCES

[1] J. H. McDermott, M. Schemitsch, and E. P. Simoncelli, "Summary statistics in auditory perception," *Nature Neuroscience*, vol. 16, pp. 493–498, Apr 2013.

[2] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

[3] J. Martínez-Más, A. Bueno-Crespo, S. Khazendar, M. Remezal-Solano, J.-P. Martínez-Cendán, S. Jassim, H. Du, H. Al Assam, T. Bourne, and D. Timmerman, "Evaluation of machine learning methods with fourier transform features for classifying ovarian tumors based on ultrasound images," *PLoS One*, vol. 14, no. 7, p. e0219388, 2019.

[4] A. Elbir, H. O. İlhan, G. Serbes, and N. Aydın, "Short time fourier transform based music genre classification," in *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, pp. 1–4, IEEE, 2018.

[5] K. M. M. Prabhu, *Window Functions and Their Applications in Signal Processing*. Taylor & Francis, 2014.

[6] R. N. Bracewell, "The fourier transform," *Scientific American*, vol. 260, no. 6, pp. 86–95, 1989.

[7] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertainty—i," *Bell System Technical Journal*, vol. 40, no. 1, pp. 43–63, 1961.

[8] M. J. Prerau, R. E. Brown, M. T. Bianchi, J. M. Ellenbogen, and P. L. Purdon, "Sleep neurophysiological dynamics through the lens of multitaper spectral analysis," *Physiology (Bethesda)*, vol. 32, pp. 60–92, Jan. 2017.

[9] M. Prerau, R. Brown, M. Bianchi, J. Ellenbogen, and P. Purdon, "Sleep neurophysiological dynamics through the lens of multitaper spectral analysis," *Physiology*, vol. 32, pp. 60–92, Jan. 2017. Funding Information: This work was supported by VA and by National Institutes of Health (NIH) Grants R21 NS-093000 (R.E.B.), R01 MH-039683 and HL-095491(R.E.B), National Institute Of Neurological Disorders And Stroke Grant R01 NS-096177 (M.J.P.), and by NIH New Innovator Award DP2-OD006454 (P.L.P.). Publisher Copyright: © 2017 Int. Union Physiol. Sci./Am. Physiol. Soc.

[10] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[11] H. Zou and T. Hastie, "Regression shrinkage and selection via the elastic net, with applications to microarrays," *JR Stat Soc Ser B*, vol. 67, pp. 301–20, 2003.

[12] C. T. Ross and K. M. Jinks, "Doj and drones: Protection: Policy, and enforcement," *Dep't of Just. J. Fed. L. & Prac.*, vol. 69, p. 275, 2021.